

IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF PENNSYLVANIA

UNITED STATES OF AMERICA, : Civil No. 1:14-cv-1474

Plaintiff,

v.

**COMMONWEALTH OF
PENNSYLVANIA and
PENNSYLVANIA STATE POLICE,**

Defendants.

: **Judge Sylvia H. Rambo**

M E M O R A N D U M

In this employment discrimination action, the United States of America sued the Commonwealth of Pennsylvania and the Pennsylvania State Police, alleging that the Commonwealth of Pennsylvania engaged in a “pattern or practice” of unlawful discrimination against females applying for employment as entry level troopers with the Pennsylvania State Police, resulting in disparate impact in violation of Title VII of the Civil Rights Act of 1964. Specifically, the United States of America challenges the Commonwealth of Pennsylvania and the Pennsylvania State Police’s use of the 2003 Physical Readiness Test and the 2009 Physical Readiness Test to screen and select applicants for trooper positions. Presently before the court are cross-motions for summary judgment. For the reasons stated herein, the court will grant Plaintiff’s motion for partial summary judgment and deny Defendants’ motion for summary judgment.

I. Background

In considering the instant motions, the court relied on the uncontested facts, or where the facts were disputed, viewed the facts and deduced all reasonable inferences therefrom in the light most favorable to the nonmoving party.

A. Facts

The Commonwealth of Pennsylvania (“Commonwealth”) and the Pennsylvania State Police (“PSP”) (collectively “Defendants”) are public employers created by the laws of the Commonwealth of Pennsylvania. Defendants are also persons within the meaning of 42 U.S.C. § 2000e(a); specifically, the Commonwealth is a state government and the PSP is a governmental agency. The parties defined certain terms in their statements of facts, which the court will adopt for ease of reference. “Trooper” refers to the entry-level position for uniformed officers with the PSP, including their time spent as a cadet at the Academy. “Academy” refers to the six-month training program that troopers must complete successfully in order to serve as uniformed officers for PSP. “Physical Readiness Test” or “PRT” refers to a battery of physical fitness events used by Defendants as part of the selection process for the PSP entry-level troopers.¹ “2003 PRT” refers to

¹ It appears that Defendants previously used the term “physical fitness test,” rather than “physical readiness test” to refer to the physical portion of the trooper selection process. (*Compare Docs. 80-1, p. 4 of 24 and 80-4, p. 4 of 13 with Doc. 80-2, p. 4 of 10.*) The court will use “physical readiness test” or “PRT” to reference the physical portion of the trooper selection process throughout this memorandum and accompanying order.

the PRT that the PSP used from May 2003 until April 2009. Finally, “2009 PRT” refers to the PRT that the PSP has used since October 2010.

The PSP provides services including criminal investigations, support of municipal police forces, traffic enforcement, and public safety throughout the entire Commonwealth of Pennsylvania. A trooper’s essential job functions include: making arrests; preparing investigative and other reports; exercising independent judgment; operating law enforcement vehicles including in emergency situations and hazardous conditions; communicating effectively and coherently with the public and when using radio channels; gathering and interpreting information for investigations; pursuing fleeing suspects which could involve lifting, climbing, jumping, and balancing; loading, unloading, aiming and firing different types of firearms in a variety of situations; performing searches of people, vehicles, buildings, and large areas; conducting surveillance; engaging in patrol functions; demonstrating effective communication skills and proper demeanor in court and other formal settings; detecting and collecting evidence; enduring verbal and mental abuse when in an antagonistic environment; performing rescue functions; processing and transporting suspects, prisoners, and committed mental patients; utilizing and operating emergency equipment; and reading and comprehending legal and non-legal documents. (Doc. 85-5, pp. 2-3.)

PSP trooper applicants must pass all steps of the selection process and then complete a six-month training academy. All residents of the Commonwealth of Pennsylvania between the ages of twenty and forty years old who meet the PSP's education requirements are eligible to apply to become a PSP trooper. Since 2003, the selection process for troopers has included the following steps: (1) an application screening; (2) a written examination; (3) an oral examination; (4) ranking on an eligibility list; (5) a physical fitness or physical readiness test; (6) a polygraph examination; (7) a background investigation; and (8) medical and psychological processing. Applicants that successfully complete the selection process receive an offer to train as a trooper at the PSP Academy.

Because the physical tasks a trooper performs involve the trooper, other troopers, and the public, the PSP seeks to determine which applicants are likely to possess the ability to perform these physical job functions through its application process. The PSP purportedly uses the PRT as a mechanism for testing the ability of trooper applicants to perform the physical job functions of the trooper position. Prior to May 2003, Defendants used a physical fitness test that had different raw cutoff scores based on age and gender. (Doc. 80-4, p. 4 of 13.) In 2001, Fitness Intervention Technologies ("FIT") was hired to identify job-related physical fitness tests, standards and programs for the PSP. In 1999, FIT collected data in an abbreviated study regarding the physical tasks of the PSP trooper job through a

“job task analysis.” FIT used this 1999 data in its 2001 study, which compared data about the physical job tasks of PSP troopers to similar data from other police agencies.² Following FIT’s study, the PSP began administering the 2003 PRT to applicants between May 2003 and April 2009. The events and cutoff scores for the 2003 PRT are:

2003 PRT	
Event	Cutoff Scores
300 meter run	67 seconds
Sit-ups	30 repetitions in one minute
Push-ups	13 repetitions in one minute ³
Vertical jump	14 inches
1.5 mile run	16 minutes 54 seconds

Prior to March 2006, the PSP contracted with FIT to conduct a validation study, a study specific to the PSP without using standards from another law enforcement agency. FIT provided its report, “Physical Fitness Standards Validation for the Pennsylvania State Police,” to the PSP in March 2006. In November 2009, the PSP began administering a new physical readiness test, which implemented the standards proposed in FIT’s report. The administration of the

² This type of study is called a transportability study.

³ There is conflicting evidence regarding whether the push-ups were timed. (*Compare Doc. 80-1, p. 4 of 24, with Doc. 85-9, p. 2 of 3.*)

2009 PRT being challenged in this litigation began on October 17, 2010 and continues through today.⁴ The events and cutoff scores for the 2009 PRT are:

2009 PRT	
Event	Cutoff Scores
Vertical jump	14 inches in 3 attempts
Illinois agility run	23.5 seconds
300 meter run	77 seconds
Push-ups	13 repetitions
1.5 mile run	17 minutes 48 seconds

Both the 2003 PRT and the 2009 PRT have unitary cutoff scores that are not age or gender based. If an applicant failed any event in either the 2003 or 2009 PRT, then he or she failed the entire PRT. From 2003 through 2008, applicants who failed the 2003 PRT were permitted to re-take the test once without re-applying. However, this practice was discontinued around February 2008.

Defendants compiled and provided data to the United States of America (“Plaintiff” or “United States”) for all applicants who took the 2003 and/or 2009 PRT between May 2003 and February 2015 (hereinafter “PRT database”), and provided a memorandum detailing the protocol followed to create the database (hereinafter “PRT database protocol”). The PRT database contains identifying information for all applicants who signed up to take the 2003 PRT and/or 2009 PRT, including their applicant identification number, first and last name, gender, birth date, application year, and the last four digits of their Social Security number.

⁴ It should be noted that the PSP did not administer the Illinois agility run from October 17, 2010 through May 8, 2011. (Doc. 80-3, p. 4 of 15.)

Additionally, the PRT database specifies the dates on which each applicant was scheduled to take the PRT, whether he or she started the PRT, and the outcomes on the PRT. It further includes records of applicants who were scheduled to take a PRT but did not actually take the PRT on the scheduled date. The parties' statistical experts relied on the PRT dataset in preparing their expert reports.

B. Procedural History⁵

Plaintiff initiated this action by filing a complaint in the Middle District of Pennsylvania on July 29, 2014. (Doc. 1.) In its complaint, the United States asserts that Defendants have violated Title VII of the Civil Rights Act of 1964, as amended, 42 U.S.C. §§ 2000e *et seq.* (“Title VII”). (*Id.*) More specifically, the United States alleges that Defendants’ use of physical fitness tests to screen and select applicants for entry-level PSP trooper positions has an unlawful disparate impact against female applicants, who have passed the physical fitness tests—and thus been hired—at significantly lower rates than male applicants. (*Id.* at ¶¶ 20-21, 23, 25, 27(b).) On November 10, 2014, Defendants filed a motion to dismiss primarily arguing that Title VII’s Section 707(a), pursuant to which the United States filed this suit, only confers authority upon the United States to bring claims for disparate treatment, and, therefore, the court lacks subject matter jurisdiction over Plaintiff’s claim for disparate impact. (Doc. 26.) Once the motion was ripe,

⁵ The court granted the parties’ request to bifurcate this matter into liability and relief phases on February 9, 2015. (Docs. 38-39.) The liability phase is currently being litigated before the court.

the court issued a memorandum and order on May 21, 2015 denying Defendants' motion to dismiss. (Docs. 45-46); *United States v. Pennsylvania*, 110 F. Supp. 3d 544 (M.D. Pa. 2015). Defendants subsequently answered the complaint on June 18, 2015. (Doc. 52.)

The United States filed a motion for partial summary judgment along with a supporting brief, declaration, and statement of facts on August 19, 2016. (Docs. 79-82.) That same day, Defendants filed their motion for summary judgment, brief in support, and statement of facts. (Docs. 83-85.) Both motions have been fully briefed (Docs. 87-95, 99-100, 106, 109, 116-117) and are ripe for disposition.

II. Legal Standard

Federal Rule of Civil Procedure 56 sets forth the standard and procedures for the grant of summary judgment. Rule 56(a) provides that “[t]he court shall grant summary judgment if the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to summary judgment as a matter of law.” Fed. R. Civ. P. 56(a); *see also Celotex Corp. v. Catrett*, 477 U.S. 317, 322-323 (1986). A factual dispute is “material” if it might affect the outcome of the suit under the applicable substantive law, and is “genuine” only if there is a sufficient evidentiary basis that would allow a reasonable fact-finder to return a verdict for the non-moving party. *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 248 (1986). When evaluating a motion for summary judgment, a court “must view the facts in

the light most favorable to the non-moving party” and draw all reasonable inferences in favor of the same. *Hugh v. Butler Cnty. Family YMCA*, 418 F.3d 265, 267 (3d Cir. 2005).

The moving party bears the initial burden of demonstrating the absence of a disputed issue of material fact. *See Celotex*, 477 U.S. at 324. “Once the moving party points to evidence demonstrating no issue of material fact exists, the non-moving party has the duty to set forth specific facts showing that a genuine issue of material fact exists and that a reasonable factfinder could rule in its favor.” *Azur v. Chase Bank, USA, Nat'l Ass'n*, 601 F.3d 212, 216 (3d Cir. 2010). The non-moving party may not simply sit back and rest on the allegations in its complaint; instead, it must “go beyond the pleadings and by [its] own affidavits, or by the depositions, answers to interrogatories, and admissions on file, designate specific facts showing that there is a genuine issue for trial.” *Celotex*, 477 U.S. at 324 (internal quotations omitted); *see also Saldana v. Kmart Corp.*, 260 F.3d 228, 232 (3d Cir. 2001). Summary judgment should be granted where a party “fails to make a showing sufficient to establish the existence of an element essential to that party’s case, and on which that party will bear the burden at trial.” *Celotex*, 477 U.S. at 322-23. “Such affirmative evidence – regardless of whether it is direct or circumstantial – must amount to more than a scintilla, but may amount to less (in the evaluation of

the court) than a preponderance.” *Saldana*, 260 F.3d at 232 (quoting *Williams v. Borough of West Chester*, 891 F.2d 458, 460-61 (3d Cir. 1989)).

III. Discussion

Title VII endeavors to ensure that workplaces are free of discrimination and that employment decisions are made based on qualifications rather than on extraneous factors such as race or gender. *NAACP v. N. Hudson Reg'l Fire & Rescue*, 665 F.3d 464, 476 (3d Cir. 2011) (citing *Ricci v. DeStefano*, 557 U.S. 557, 580 (2009)). To that end, “Title VII makes it illegal for an employer to ‘limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities . . . because of such individual’s race, color, religion, sex, or national origin’” *Id.* (quoting 42 U.S.C. § 2000e-2(a)(2)). Title VII prohibits both intentional discrimination, known as disparate treatment discrimination, as well as, in some cases, practices that are not intended to discriminate but in fact have a disproportionately adverse effect on minorities, known as disparate impact discrimination. *Ricci*, 557 U.S. at 577.

Disparate impact discrimination, the type of discrimination alleged in this action, is a brand of “unintentional discrimination,” whereby “employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another” are prohibited. *Int'l Bhd. of*

Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977). Although the employer acts without a deliberately discriminatory motive, its actions are functionally equivalent to intentional discrimination. *Ladd v. Boeing Co.*, 463 F. Supp. 2d 516, 521 (E.D. Pa. 2006) (citing *Watson v. Ft. Worth Bank & Trust*, 487 U.S. 977, 987 (1988)). Consequently, evidence submitted in disparate impact cases tends to focus on statistical disparities rather than specific incidents, and on competing explanations for those disparities. *Id.*

The Civil Rights Act of 1964 did not include an express prohibition on policies or practices that result in a disparate impact. *Ricci*, 557 U.S. at 578. Rather, the disparate impact theory of discrimination under Title VII was judicially created by the Supreme Court in 1971 in the seminal case of *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). Recognizing that Congress' objective in enacting Title VII was to achieve equality and remove barriers in employment, the Court held that Title VII was meant not only to proscribe overt discrimination, but also to prevent "practices that are fair in form, but discriminatory in operation." *Griggs*, 401 U.S. at 429-31.

Twenty years after *Griggs*, Congress codified the prohibition on disparate impact discrimination in the Civil Rights Act of 1991. *Ricci*, 557 U.S. at 578; see *N. Hudson Reg'l Fire & Rescue*, 665 F.3d at 477 n.9 (noting that the Civil Rights Act of 1991 reversed a Supreme Court case that had weakened the standard for

finding employment practices not a violation of Title VII). Under the statute, the plaintiff must “demonstrate[] that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.” 42 U.S.C. § 2000-e2(k)(1)(A)(i). The employer may then defend against the plaintiff’s *prima facie* showing of disparate impact only by demonstrating that the practice is “job related for the position in question and consistent with business necessity.” *Id.* Even if the employer meets that burden, however, the plaintiff may still succeed “by showing that alternative practices would have less discriminatory effects while ensuring that candidates are duly qualified.” *N. Hudson Reg’l Fire & Rescue*, 665 F.3d at 477 (citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii),(C)).

Defendants move for summary judgment on three bases: (1) that disparate impact claims are not cognizable under 42 U.S.C. § 2000e-6(a)⁶ because of *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507 (2015); (2) that Title VII on its face prohibits different employment standards based on sex; and (3) that the United States fails the last element of its burden by suggesting the adoption of gender-based fitness tests as an alternative selection device to the PRT. (Doc. 84.) In contrast, Plaintiff argues in favor of partial summary judgment that it has satisfied the first prong of the Title

⁶ Consistent with the parties’ previous filings and this court’s May 21, 2015 memorandum, the court will refer to this provision by its Statutes at Large section number, Section 707(a).

VII burden-shifting framework by showing that women pass the 2003 PRT and 2009 PRT at statistically significantly lower rates than men. (Doc. 82.) The court will address each argument in turn.

A. The United States' Authority to Assert a Disparate Impact Claim Under Section 707(a)

Defendants' principal argument is that Title VII disparate impact claims are not within the jurisdiction of this court under Section 707(a) because of the Supreme Court's recent opinion in *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.* (Doc. 84, pp. 12-22 of 36.) According to Defendants, *Inclusive Communities* sets forth the proper analysis to "interpret statutes that purport to create a cause of action for disparate impact." (*Id.* at 12-13 of 36.) Defendants present a lengthy argument relying on *Inclusive Communities* to reinterpret Section 707 and conclude that it "cannot validly be interpreted to allow disparate impact claims," thus, the court lacks jurisdiction. (*Id.* at 20 of 36.) They further argue that all decisions issued prior to *Inclusive Communities* are not binding and the court has no basis to rely on anything other than *Inclusive Communities*. (*Id.* at 21-22 of 36.)

The Commonwealth and the PSP previously presented this jurisdictional theory in their motion to dismiss. (Docs. 26 & 31.) There, Defendants set forth a nearly identical statutory interpretation argument and relied, in part, on *Inclusive Communities* in the motion to dismiss even though the Supreme Court had only

granted certiorari at that time.⁷ (Doc. 31, pp. 17-24 of 30.) In its May 21, 2015 memorandum denying Defendant's motion to dismiss and holding that the court had subject matter jurisdiction under Section 707, the court stated:

Defendants further argue that the court should be guided by the Supreme Court's decision to grant certiorari in a case brought pursuant to the Fair Housing Act ("FHA"), 42 U.S.C. §§ 3601, et seq., *Texas Department of Housing and Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 46 (2014). The question presented in *Inclusive Communities* is whether, based on the text of the FHA, disparate impact claims are cognizable under the FHA. Brief for Petitioner at 5, *Inclusive Communities*, 135 S. Ct. 36 (2014) (No. 13-1371), 2014 WL 1989121 at *1. Because there is no question that disparate impact claims are cognizable under Title VII, the grant of certiorari in *Inclusive Communities* is irrelevant. Nothing in *Inclusive Communities* casts any doubt on the continued validity of disparate impact cases under Title VII.

United States v. Pennsylvania, 110 F. Supp. 3d 544, 552 n.6 (M.D. Pa. 2015).

To begin with, Defendants' contention that their statutory interpretation argument is a proper motion for summary judgment under Federal Rule of Civil Procedure 56 is disingenuous. (See Doc. 95, p. 5 of 23.) As Plaintiff indicates, Defendants essentially request that the court revisit and overturn its May 21, 2015 ruling due to the final decision in *Inclusive Communities*. (See Doc. 90, pp. 10-11 of 31.) The law of the case doctrine gives the court "the power to revisit prior

⁷ On June 25, 2015, approximately one month after this court issued its memorandum and order regarding Defendants' motion to dismiss, the Supreme Court affirmed and remanded *Inclusive Communities*. *Inclusive Cmty.*, 135 S. Ct. at 2512.

decisions of its own or of a coordinate court in any circumstance, although as a rule courts should be loathe to do so in the absence of extraordinary circumstances.” *Christianson v. Colt Indus. Operating Corp.*, 486 U.S. 800, 817 (1988). This doctrine is a “rule of discretion rather than a limit on authority.” *Schneyder v. Smith*, 653 F.3d 313, 331 (3d Cir. 2011). The Third Circuit articulated three examples of extraordinary circumstances when the law of the case doctrine does not discourage revisiting a previous order: “(1) new evidence is available; (2) a supervening new law has been announced; or (3) the earlier decision was clearly erroneous and would create manifest injustice.” *Id.* at 331-32.

While the court recognizes that *Inclusive Communities* is new law, it is not supervening new law that affects the court’s jurisdiction over Title VII disparate impact cases. *Inclusive Communities* held that disparate impact claims were cognizable under the Fair Housing Act. *Inclusive Cmty.*, 135 S. Ct. at 2525. In coming to this conclusion, the Supreme Court considered the history of the two antidiscrimination statutes that preceded the Fair Housing Act, Title VII and the Age Discrimination in Employment Act. *Id.* at 2516-25. The Supreme Court’s decision in no way changed the legal framework for disparate impact claims under Title VII or the Age Discrimination in Employment Act. Rather, it expanded the disparate impact theory with regards to the Fair Housing Act. Hence, nothing in the Supreme Court’s opinion changes or undercuts the court’s finding that “the United

States is authorized to bring a pattern or practice claim of disparate impact discrimination under 707(a).” *United States*, 110 F. Supp. 3d at 551.

Accordingly, the court finds that *Inclusive Communities* does not provide supervening new law in the Title VII disparate impact context to warrant revisiting the court’s May 21, 2015 decision, and will deny Defendants’ motion for summary judgment in this regard.

B. Prima Facie Showing of Disparate Impact Discrimination

In its motion for partial summary judgment, the United States argues that it can establish a *prima facie* case of disparate impact discrimination under prong one of Title VII’s three-prong burden shifting test. (Doc. 82.) Specifically, Plaintiff avers that it can prove through statistical evidence that Defendants’ use of the 2003 PRT and 2009 PRT have a disparate impact on women. (*Id.* at 13-28 of 31.)

To establish a *prima facie* case of disparate impact, a plaintiff must identify “the specific employment practice that is challenged”⁸ and show that the employment practice “causes a disparate impact on the basis of race, color, religion, sex, or national origin.” *Watson*, 487 U.S. at 994; 42 U.S.C. § 2000e-2(k)(1)(A)(i). To show causation, a plaintiff must provide “statistical evidence of a kind and degree sufficient to show that the practice in question has caused exclusion of applicants for jobs or promotions because of their membership in a

⁸ The parties do not dispute that the United States clearly identifies the employment practices that it is challenging, the 2003 PRT and 2009 PRT.

protected group.” *Watson*, 487 U.S. at 944; *see also Stagi v. Nat'l R.R. Passenger Corp.*, 391 F. App’x 133, 136 (3d Cir. 2010); *EEOC v. Greyhound Lines*, 635 F.2d 188, 193 (3d Cir. 1980).

There is no strict mathematical formula for courts to apply to determine whether a plaintiff has established a *prima facie* case, but the Supreme Court has “consistently stressed that statistical disparities must be sufficiently substantial that they raise such an inference of causation.” *Watson*, 487 U.S. at 994-95. The disparity needs to be sufficiently large so that “it is highly unlikely to have occurred at random . . . using one of several tests of statistical significance.” *Stagi*, 635 F.2d at 144. Courts, including the Third Circuit, typically measure statistical significance by probability levels (“p-values”) and standard deviation.⁹ *Id.* at 137, 140. Although there is no precise threshold, “a finding of statistical significance with a probability level at or below 0.05, or at 2 to 3 standard deviations or greater, will typically be sufficient” to meet plaintiff’s burden. *Id.* at 140 (citing *Castaneda v. Partida*, 460 U.S. 482, 496 n.17 (1977) (“As a general rule for . . . large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the [result] was random would be suspect to a social scientist.”)); *see also Waisom v. Port Auth.*,

⁹ P-values are “the probability that the observed disparity is random—the result of chance fluctuation or distribution,” while standard deviation is “a unit of measurement that allows statisticians to measure all types of disparities in common terms.” *Stagi*, 635 F.2d at 137.

948 F.2d 1370, 1376 (2d Cir. 1991) (“Social scientists consider a finding of two standard deviations significant, meaning there is about one chance in 20 that the explanation for a deviation could be random and the deviation must be accounted for by some factor other than chance.”). Using statistical significance rather than another method allows “a fact-finder to be confident that the relationship between some rule or policy and some set of disparate impact results was not the product of chance.” *Stagi*, 391 F. App’x at 145 (citing *Watson*, 487 U.S. at 994-95).

Some courts have used the Equal Employment Opportunity Commission’s (“EEOC”) Uniform Guidelines on Employee Selection Procedures “80 percent rule” to determine whether a plaintiff has established a *prima facie* disparate impact case. *Id.* at 138. However, the “‘80 percent rule’ or ‘four-fifths rule’ has come under substantial criticism, and has not been particularly persuasive, at least as a prerequisite for making out a *prima facie* disparate impact case.” *Id.*; see *Watson*, 487 U.S. at 995 n.3 (“This enforcement standard has been criticized on technical grounds . . . and it has not provided more than a rule of thumb for the courts.”). Thus, to establish a *prima facie* disparate impact case, a plaintiff does not need to show that the disparate impact ratio satisfies the EEOC’s 80 percent rule. *Stagi*, 391 F. App’x at 144.

The United States presents Dr. Janice Madden, a labor economist and statistical expert, in support of its *prima facie* case. (See Doc. 80-11.) Defendants

provide three experts to rebut Dr. Madden: Dr. Brice Stone, a statistical expert; Dr. William Fairley, also a statistical expert; and Dr. Palmer Morrel-Samuels, an experimental social psychologist. (*See* Docs. 80-8, 80-9, & 80-10.)

1. Dr. Madden's Analysis

Dr. Madden compared the pass rates of the 2003 PRT and 2009 PRT of female applicants to the pass rates of male applicants. (Doc. 80-11, p. 6 of 37.) She counted an individual as taking the 2003 or 2009 PRT if the individual was described by the PRT database as: (1) having an outcome recorded, such as pass or fail; and (2) having quit or withdrawn from the PRT after starting it. (*Id.* at 7 of 37 n.2.) Dr. Madden utilized the multiple pools exact test, a widely used statistical significance test, to determine whether any observed disparities in the passing rates of men and women on the 2003 PRT or the 2009 PRT could be attributed to random chance. (*Id.* at 7-8 of 37.) Using the multiple pools exact test, Dr. Madden analyzed the gender differences in pass rates for the 2003 and 2009 PRTs in three ways: (1) *each time* an applicant took a PRT (Table 1a); (2) each applicants *first time* taking a PRT (Table 1b); and (3) each applicant's *last time* taking a PRT (Table 1c). (*Id.* at 9-10, 17 of 37.) Dr. Madden opines that using the results for *each time* an applicant took a PRT, as detailed in Table 1a of her report, is best for measuring whether there is a gender disparity in PRT outcomes because it uses all the information on test takers, including the results for applicants who took the

PRT more than one time. (*Id.* at 9 of 37.) For each of the data sets, Dr. Madden calculated an overall p-value for the time period in which the 2003 PRT and 2009 PRT were administered and then converted the p-values into equivalent standard deviations. (*Id.* at 8 of 37.)¹⁰

Dr. Madden's primary analysis as described in Table 1a, and recreated below, details the results for *each time* an applicant took a PRT. (*Id.* at 17 of 37.)

Gender differences in passing the PRT based on all PRT attempts									
Years test was administered	Women			Men			Standard deviation equivalents: gender difference in pass rates	Women's pass rate as percentage of men's pass rate	Number of additional women who would have passed with equivalent pass rates by gender
	Number taking test	Number passing test	Pass rate	Number taking test	Number passing test	Pass rate			
2003-2009	578	318	55%	4,735	4,180	88%	>7.0	62%	151.8
2010-2015	572	419	73%	5,363	5241	98%	>7.0	75%	123.0

¹⁰ Dr. Madden did not report units of standard deviations greater than seven because she opines that there is no meaningful distinction after seven units are reached. (Doc. 83-13, p. 33 of 77.)

As detailed in the recreated tables below, Dr. Madden also analyzed the passing rates for each applicants *first time* taking a PRT (Table 1b), and each applicant's *last time* taking a PRT (Table 1c). (*Id.* at 17 of 37.)

Gender differences in passing the PRT based on first PRT attempt									
Years test was administered	Women			Men			Standard deviation equivalents: gender difference in pass rates	Women's pass rate as percentage of men's pass rate	Number of additional women who would have passed with equivalent pass rates by gender
	Number taking test	Number passing test	Pass rate	Number taking test	Number passing test	Pass rate			
2003-2009	470	268	57%	4,275	3,802	89%	>7.0	64%	122.6
2010-2015	525	381	73%	4,901	4,787	98%	>7.0	74%	115.1

Years test was administered	Gender differences in passing the PRT based on last PRT attempt								
	Women			Men			Standard deviation equivalents: gender difference in pass rates	Women's pass rate as percentage of men's pass rate	Number of additional women who would have passed with equivalent pass rates by gender
	Number taking test	Number passing test	Pass rate	Number taking test	Number passing test	Pass rate			
2003-2009	470	294	63%	4,275	3,927	92%	>7.0	68%	111.4
2010-2015	525	390	74%	4,901	4,797	98%	>7.0	76%	108.0

These additional analyses resulted in the same conclusion as the primary analysis: pass rates exceeded seven standard deviations, thus, women passed the 2003 PRT and 2009 PRT at statistically significantly lower rates than men. (*Id.* at 12-13, 17 of 37.)

2. Dr. Stone's Analysis

Dr. Stone, one of Defendants' experts, analyzed the difference in pass rates between men and women on the 2003 PRT and the 2009 PRT by studying applicant's "last chronological record" in the PRT database for each year the 2003 or the 2009 PRT was administered.¹¹ (Doc. 80-8, p. 5 of 36.) In his analysis, Dr. Stone removed "duplicate records for each individual," that is, any instances where an applicant attempted the PRT more than once with the exception of the applicant's last chronological record. (*Id.* at 5 of 36.)

Dr. Stone states in his report: "It should be noted that the distribution of PRT outcomes for males versus females (presented in Table 3) [for the 2003 to 2014

¹¹ Dr. Stone limited his analysis to those PRTs administered between 2003 and 2014. (Doc. 80-8, p. 5 of 36.)

time period] were statistically significantly different from one another.” (*Id.* at 7 of 36.) In his deposition, he testified that the p-value for the difference in male and female pass rates for the 2003 PRT and 2009 PRT is 0.1 or lower. (Doc. 80-15, p. 18 of 53.) He reports that from 2003 to 2008, the average pass rate for males was 96.68 percent, while the pass rate for females was 86.44 percent, which is a statistically significant difference. (Doc. 80-8, p. 8 of 36.) He concludes that the female pass rate is compliant with the 4/5ths rule as compared to the male pass rate, exhibiting a ratio of 0.885. (*Id.*) For the 2010 to 2014 time period, 98.51 percent of the male applicants passed the PRT with 78.63 percent of the female applicants passing, a statistically significant difference. (*Id.* at 9 of 36.) This ratio of 0.798 is “slightly less compliant” with the 4/5ths rule as compared to the male pass rate. (*Id.*) Dr. Stone opines that “rounding this ratio to the first decimal results in a 0.8 ratio, [and is] compliant with the 4/5ths rule.” (*Id.*)

3. Dr. Fairley’s Analysis

As Defendants’ rebuttal expert, Dr. Fairley focuses his attention on critiquing Dr. Madden’s analysis but does not provide any results of his analysis of the PRT database or his early statistical significance testing other than to say that there is a difference in the pass rates. (*Id.* at 4-13 of 28.) His criticisms include that Dr. Madden’s report is exaggerated, contains false precision of probabilities, and

assumes a particular and simplified theoretical model of probabilities. (Doc. 80-9, pp. 4-12 of 28; Doc. 88-2.)

As for providing his own statistical analysis, Dr. Fairley declined to do so, indicating instead that “[t]here is a difference in pass rates between men and women in each of the test date periods.” (Doc. 80-9, p. 4 of 28.) Dr. Fairley did not calculate a p-value for the difference in pass rates between men and women, indicating that it was “not necessary” because there is “[s]o much data. It’s a foregone conclusion that the p-value is going to be less than .01.” (Doc. 80-16, p. 15 of 56.) Thus, without doing any calculation, he agreed in his deposition that the difference between men and women passing the 2003 and 2009 PRT was statistically significant. (*Id.*) Dr. Fairley further testified that he did not formally apply any measures of statistical significance, *i.e.* make a calculation, such as a p-value or standard deviation, to the PRT database. (*Id.*) However, prior to the creation of the PRT database, Dr. Fairley used a collection of files from the PSP to run early statistical significance testing and found that there was a statistically significant difference between pass rates of men and women on the 2003 PRT and 2009 PRT. (*Id.* at 24 of 56.)

4. Dr. Morrel-Samuels’ Analysis

Dr. Morrel-Samuels, Defendants’ second rebuttal expert, describes four flaws with Dr. Madden’s report. (Doc. 80-10, pp. 2-14.) He first asserts that

“counting standard deviations based on an expectation of exact parity is misleading on methodological grounds, and unjustified on statutory grounds.” (*Id.* at 3 of 14.) Next, Dr. Morrel-Samuels opines that Dr. Madden’s pooling method is flawed because it excludes eleven percent of the test administrations from the dataset, and “pooling gives each day an equal ‘weight’ regardless of numbers and proportions.” (*Id.* at 4 of 14.) Although Dr. Madden’s used the multiple pools exact test, Dr. Morrel-Samuels states that the Fishers Exact Test is “ill-suited for analyzing the wide variety of circumstances that we encounter in the PSP dataset.” (*Id.*) Third, Dr. Morrel-Samuels asserts that Dr. Madden’s “error of identifying ‘a shortfall of 3.8 women’ reveals the folly of its computational method” as 3.8 women cannot be harmed by a PRT. (*Id.* at 5-6 of 14.) Lastly, the most serious flaw in Dr. Madden’s report is the inability to measure the impact of nondiscriminatory factors such as age, weight, height, practice taking the test, and military service. (*Id.* at 6 of 14.)

5. Comparison of Expert Reports and Analysis

While the parties’ respective experts disagree on the appropriate method of statistical analysis, Dr. Madden, Dr. Stone, and Dr. Fairley agree that women pass the 2003 PRT and 2009 PRT at statistically significantly lower rates than men; specifically, at a p-value of 0.1 or lower. Dr. Stone, Defendants’ primary statistics expert, frames his analysis in terms of the 80 percent or four-fifths rule. Defendants argue that the United States is bound to follow this rule, and the court should adopt

Dr. Stone's analysis, due to the language in Plaintiff's complaint.¹² (Doc. 87, pp. 10-11 of 28.) The court will not entertain Dr. Stone's conclusion that the four-fifths rule is satisfied and, therefore, Plaintiff has not met its *prima facie* burden. As discussed earlier, *supra* section III(B), the four-fifths rule is a non-statistical standard that courts within the Third Circuit no longer follow. *See Stagi*, 391 F. App'x at 144.

As Dr. Madden points out in her rebuttal report, Dr. Stone's analysis removes many PRT test takers by restricting his analysis to the last PRT record for each applicant and by removing all last PRT records that occurred in 2015. (Doc. 80-8, p. 5 of 36; Doc. 80-12, p. 5 of 21.) According to Dr. Madden's analysis, Dr. Stone eliminated 65 PRT outcomes because the last record occurred in 2015. (Doc. 80-12, pp. 5 & 19 of 21.) He also deleted 774 applicants who took the PRT but whose last record did not include any PRT results. (*Id.*) Dr. Madden points to several examples of records Dr. Stone erroneously eliminated without any stated reason:

- Tonya Gordon who failed the PRT on August 22, 2008 and did not report for the PRT on December 18, 2008,
- Laura Mykytok who failed the PRT on May 23, 2006 and did not report for the PRT on August 24, 2006,
- Amanda Dunmyer who failed the PRT on June 9, 2008 and did not report for the PRT on August 20, 2008.

¹² At several points in the complaint, the United States refers to the difference in pass rates between male and female applicants as statistically significant, and that "the female pass rate is less than 80% of the male pass rate." (Doc. 1, ¶¶ 14-15, 18-19.)

(*Id.* at 5 of 21.) If Dr. Stone had included those 839 records, the four-fifths rule would not have been satisfied. (*Id.* at 5-7 of 21.)

Because Dr. Stone eliminated the 2015 PRT records from his analysis, Dr. Madden repeated her initial analysis for the 2009 PRT without the PRT outcomes from 2015. (*Id.* at 7 & 20 of 21.) Her revised results excluding 2015 PRT outcomes were similar to her original results: “The 76% average pass rate for women on the [2009 PRT] as administered between 2010 and 2014 is more than seven standard deviations lower than the 98% average pass rate for men taking the exam on the same day.” (*Id.* at 7.)

All three of Defendants’ experts agree that there is statistically significant disparity between men and women in passing rates on the 2003 and 2009 PRTs. However, other than Dr. Stone’s four-fifths rule analysis, none of the experts provide any calculations contrary to or disproving Dr. Madden’s probability values and standard deviation results. Rather, Defendants’ experts focus on discrediting Dr. Madden’s method of analysis, but do not suggest a more appropriate method with corresponding calculations. For example, Dr. Morrel-Samuels disapproves of the Fishers Exact Test but does not acknowledge that Dr. Madden used the more conservative multiple pools exact test. Further, Dr. Morrel-Samuels criticizes Dr. Madden because her statistical analysis flows from the assumption that if random chance were operating, men and women would be expected to have the same

probability of passing the PRT. Courts have noted that in order to determine what the results would be without the salient variable of gender or race, one must assume that the group will perform equally well absent discrimination. *United States v. City of New York*, 637 F. Supp. 2d 77, 96-97 (E.D.N.Y. 2009). “Statistical significance testing relies on this assumption of equality in assessing whether disparities among groups are based upon chance, or rather, upon some other fact, such as race or national origin.” *Id.* at 97 (citing *Ottaviani v. State Univ. of N.Y. at New Paltz*, 875 F.2d 365 (2nd Cir. 1989)).¹³

Defendants and their experts also argue that the United States cannot meet its *prima facie* burden because Dr. Madden’s analysis did not eliminate other possible explanations for passing rates. (Doc. 87, pp. 20-22 of 28.) However, in support of their argument, Defendants rely on numerous cases evaluating disparate treatment, rather than disparate impact. Specifically, Defendants heavily rely on *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940 (7th Cir. 1997), as “a proper

¹³ The Seventh Circuit Court of Appeals provided a helpful example:

If the relevant market is 40% African-American, for instance, one would expect 40% of hires to be African-American. . . . If the observed percentage of African-American hires is only 20%, then the statistician will compute the ‘standard deviation’ from the expected norm and indicate how likely it is that race played no part in the decision making. Two standard deviations is normally enough to show that it is extremely unlikely (that is, there is less than a 5% probability) that the disparity is due to chance, giving rise to a reasonable inference that the hiring was not race-neutral; the more standard deviations away [from zero], the less likely the factor in question played no role in the decision making process.

Adams v. Ameritech Servs., Inc., 231 F.3d 414, 424 (7th Cir. 2000).

judicial analysis” for evaluating if Dr. Madden controlled or accounted for other factors besides gender. (Doc. 87, pp. 20-22 of 28.) Reliance on *Sheehan* and other disparate treatment cases is completely misguided, however, as the standards for evaluating disparate impact and disparate treatment cases are dissimilar. At the *prima facie* stage, the United States does not have the burden of eliminating all other possible explanations for passing rates – rather, it must simply proffer evidence “of a kind and degree sufficient to show that the practice in question has caused exclusion of applicants for jobs or promotions” because of their gender. *Stagi*, 391 F. App’x at 144.

The court is satisfied that the United States has met its *prima facie* burden of disparate impact and there is no genuine issue of material fact regarding this prong of the disparate impact analysis. Dr. Madden provides a comprehensive analysis of data showing women passed the 2003 PRT and 2009 PRT at statistically significantly lower rates than men with probability values far less than 0.5 and standard deviations greater than seven. Dr. Stone, Dr. Fairley, and Dr. Morrel-Samuels do not provide convincing evidence, even when viewed in a light most favorable to Defendants, that Dr. Madden’s calculations are faulty or unreliable. Using the PRT database, Defendants’ own dataset, the United States provided sufficient statistical evidence of a kind and degree sufficient to show that the 2003 PRT and 2009 PRT has excluded women from continuing the hiring process to

become a trooper because of their gender. Accordingly, the court will grant Plaintiff's motion for partial summary judgment.

C. **Business Necessity of the PRT**

Because the United States has met its *prima facie* burden, Defendants can only defend this showing of disparate impact by demonstrating that the 2003 and 2009 PRT is “job related for the position in question and consistent with business necessity.” 42 U.S.C. § 2000e-2(k)(1)(A)(i). The Third Circuit has interpreted this to mean “that employers may not use criteria which have a discriminatory effect unless those criteria define the minimum qualifications necessary to perform the job.” *N. Hudson Reg'l Fire & Rescue*, 665 F.3d at 477 (citing *Lanning v. Se. PA. Transp. Auth.*, 181 F. 3d 478, 489 (3d Cir. 1999) (“[I]n order to show the business necessity of a discriminatory cutoff score an employer must demonstrate that its cutoff measures the minimum qualifications necessary for successful performance of the job in question.”).

The Defendants as the employer bear the burdens of production and persuasion and must provide *actual* reasons why the challenged employment practice, *i.e.*, the 2003 and 2009 PRT, is important to the trooper position. *Id.* (citations omitted). The “mere assertion of . . . conceivable bases is not sufficient.” *Id.* (quoting *NAACP v. Town of Harrison, N.J.*, 940 F. 2d 792, 804 (3d Cir. 1991)). Further, the Supreme Court has “refused to accept bare or ‘common-sense’-based

assertions of business necessity and instead required some level of empirical proof that challenged hiring criteria accurately predicts job performance.” *El v. Se. Pa. Transp. Auth.*, 479 F.3d 232, 240 (3d Cir. 2007) (referring to multiple Supreme Court opinions including *Griggs v. Duke Power*, 401 U.S. 424 (1971) and *Dothard v. Rawlinson*, 433 U.S. 321 (1977)). If an employer cannot provide compelling evidence of business necessity, the plaintiff wins simply by proving a *prima facie* case. *N. Hudson Reg'l Fire & Rescue*, 665 F.3d at 477. Conversely, if the employer can present the appropriate evidence of business necessity, the plaintiff can overcome the employer’s defense by “showing that alternative practices would have less discriminatory effects while ensuring that candidates are duly qualified.” *Id.* (citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C)).

In the case at hand, the Commonwealth and the PSP have not moved for summary judgment nor presented any compelling evidence or argument regarding the business necessity of the 2003 PRT and 2009 PRT. Rather, Defendants ignore their burden and try to direct the court’s attention to the third part of the burden shifting test. As the United States aptly points out, Defendants are putting the cart before the horse. Defendants have the burdens of production and persuasion to show the court actual reasons why the 2003 and/or 2009 PRT is job related for the trooper position and consistent with business necessity. Because Defendants have

not done so, the court need not address their arguments concerning the third prong of the disparate impact analysis. (*See* Doc. 84, pp. 22-34 of 36.)

IV. Conclusion

For the reasons stated herein, the court will grant Plaintiff's motion for partial summary judgment and deny Defendants' motion for summary judgment. An appropriate order will issue.

s/Sylvia Rambo

SYLVIA H. RAMBO
United States District Judge

Dated: October 2, 2017